
Piovesan, Walter. "Mounting a Full-Text Database Using SPIRES."
The Public-Access Computer Systems Review 1, no. 3 (1990): 83-88.

1.0 Introduction

The demand for enhanced online services has led many libraries to provide users with access to machine-readable indexes and other products in addition to the online catalogue. The proliferation of networks and the merging of two heretofore separate service bureaus--the library and computer services, has facilitated the emergence of new partnerships providing new, improved services.

This article describes how the Library and Computer Services of Simon Fraser University worked together to select and mount the GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA database on a mainframe using the SPIRES system.

2.0 Database Selection

In the summer of 1986, the Vice President for Research and Information Services at Simon Fraser University, who was responsible for both the Library and Computing Services, called together staff from both units. The Vice President had just returned from the 1986 Education Conference held at Carnegie Mellon University, and he had been impressed with the emerging new library information systems that were being demonstrated there. He requested that a working group be formed to investigate what new types of databases we could provide to the campus, such as index, encyclopedia, dictionary, and directory databases.

As Head of the Research Data Library, I was responsible for the collection and maintenance of machine-readable data for the campus community. Consequently, I was asked to head the project and to report back with a list of databases that would be feasible to load onto the campus mainframe. The databases that were identified as being suitable for the initial phase of the project were CURRENT CONTENTS, ERIC, GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA, MEDLINE, and PSYCHINFO.

A working team of Wolfgang Richter, a Database Administrator from Computing Services, and myself was formed. We were asked to load the ERIC, GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA, and PSYCHINFO databases on the campus mainframe. All of these databases were subsequently loaded.

The Database Administrator had already designed a menu-driven

user interface to a number of applications on our central mainframe: e-mail, word processing, CS Newsletter, and the exam schedule. These services were part of EASYMTS (MTS being our operating system). We decided that we would add an additional level of menus--InfoServe--which would contain an array of library-based services.

3.0 Selection of SPIRES

Prior to ordering the Grolier database, we contacted Nancy Evans of Carnegie Mellon University, who provided some key bits of information on how they had approached the task of loading the Grolier database into their STAIRS system. The main point that Ms. Evans stressed was the need for full-text indexing.

The Database Manager and myself then met to decide on which of the two campus database management systems--SPIRES or ORACLE-- we would choose to load the Grolier database into. After an examination of the pros and cons of each system, we settled on SPIRES. The main reasons for this decision were that SPIRES had: (1) the ability to easily index on individual words; (2) high-performance characteristics; (3) superior and flexible report generation capabilities; (4) the ability to easily handle large data files; and (5) superiority in handling multiple users on our IBM mainframe computer.

4.0 Characteristics of the Grolier Database

The GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA, which is approximately 170 megabytes in size, comes in the form of a single file on magnetic tape. The cost of subscribing to the database is based on size of the institution. There are quarterly updates.

5.0 Pre-Load Activities

In late 1986, we ordered a sample copy of the GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA database. The Database Administrator designed a SPIRES database definition (called a FILEDEF) and a report definition (called a FORMATS definition) for displaying search results.

The FILEDEF would allow for indexing on every word. We realized that this would make for a lengthy process in loading the full database, but we knew that if the product was to be successful with users it had to be fully indexed.

+ Page 85 +

After giving a demonstration of the Grolier database, we received approval to purchase the full database and proceed to make it available via the expanded EASYMTS service as a part of the

InfoServe menu.

Once we started to load the full database, we had to make a minor change to the existing FILEDEF. In the initial FILEDEF, each item in the database corresponded to an article; however, this proved problematic with large encyclopedia articles. The FILEDEF was modified so that we would have smaller units of information: paragraphs.

The database was indexed on four principal fields: (1) article number (this is mostly useful for the database manager and is used for checking for duplicate articles), (2) article name (3) text type (e.g., bibliographic, tables, and see also references), and (4) word (this is every word in the encyclopedia, excluding the common words like "as," "is," and "to").

6.0 Loading the Database

To ensure that any database errors were identified prior to loading the database into SPIRES, the Database Administrator wrote a series of utility programs. The programs scan the data on tape to ensure that: (1) all the fields are present, (2) fields are properly delineated, (3) there are no duplicate article numbers and that numbers be of the correct length, and (4) the information is the proper sequence as specified by the vendor. (Interested SPIRES users can contact the author to obtain copies of these utility programs, which tend to be specific to the MTS operating system.)

There were some initial problems with the database, such as errors in format and improperly delimited fields. We were able to easily identify the errors and correct them prior to loading.

Processing the database through our error checking programs added a couple of extra steps to the process, but we found that the extra time spent is well worthwhile as it saves us time in the long run. Although we found errors during the initial database load, the database has been very stable for the past two years.

+ Page 86 +

7.0 Processing Quarterly Updates

The quarterly updates for the Grolier database are processed as follows.

First, we copy the tape data to disk and run the above-mentioned checking programs, which alert us to errors that need correcting. This checking is done via utility programs specific to our MTS operating system.

Second, we correct any errors and run a FORTRAN program to convert the data into the SPIRES batch-load format. This "tags"

the database for loading into SPIRES, somewhat like adding MARC tags for loading bibliographic data into an OPAC.

Third, we batch load the data into a test subfile using the SPIBILD program. We briefly check the data with SPIRES for glaring errors, such as duplicate article numbers.

Fourth, we run a utility program that: (1) dumps out the data from the test subfile, (2) checks the main database for articles with the same name (the Grolier people do not flag updated material as such--we have to deduce it), and (3) automatically generates the appropriate set of SPIRES REMOVE and ADD commands for SPIBILD.

Finally, we run an overnight job so that SPIBILD can process the REMOVE and ADD commands generated in the previous step. We process half of the Grolier database at one time in order to reduce down time as much as possible. It takes approximately 3 hours of CPU time on our 3091 IBM mainframe to process half of the database (the elapsed clock time comes to about 14 hours). SPIRES spends most of the processing time updating the article text index, which is based on individual words used in articles.

At the time that we update the database, we insert an edition statement so that when users select the database they will know how current the information in it is.

+ Page 87 +

8.0 Reactions to the Grolier Database

During our initial investigation of the products that we wanted to offer on the InfoServe service there was some skepticism on the part of librarians who felt that students would not be able to properly search the databases and that the GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA would not meet the needs of university students. After three years of using the service and hearing from students that they really find the encyclopedia useful and use it regularly, the librarians have come to appreciate the need for self-serve reference information and are encouraging us to find other products to load, such as dictionaries. There are on average 1,200 searches per month on the GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA database.

It has also proved to be very successful with the Education department, which uses the encyclopedia in their courses on computers and information that they give to high school students. These students have no problem in using the service.

9.0 Conclusion

Using the SPIRES software, Simon Fraser University has successfully mounted the full-text GROLIER ACADEMIC AMERICAN ENCYCLOPEDIA and other databases. The encyclopedia database has

received a warm reception from the university community, and it has proven itself to be a valuable information resource.

+ Page 88 +

About the Author

Walter Piovesan
Head, Research Data Library
W.A.C. Bennett Library
Simon Fraser University
Burnaby, British Columbia, CANADA
BITNET: USERVINO@SFU.BITNET
Internet: walter_piovesan@cc.sfu.ca

The Public-Access Computer Systems Review is an electronic journal. It is sent free of charge to participants of the Public-Access Computer Systems Forum (PACS-L), a computer conference on BITNET. To join PACS-L, send an electronic mail message to LISTSERV@UHUPVM1 that says: SUBSCRIBE PACS-L First Name Last Name.

This article is Copyright (C) 1990 by Walter Piovesan. All Rights Reserved.

The Public-Access Computer Systems Review is Copyright (C) 1990 by the University Libraries, University of Houston. All Rights Reserved.

Copying is permitted for noncommercial use by computer conferences, individual scholars, and libraries. Libraries are authorized to add the journal to their collection, in electronic or printed form, at no charge. This message must appear on all copied material. All commercial use requires permission.
